SUPPLEMENTARY NOTE 2

Stage 2 peaBrain predictions were significantly and positively correlated with univariate eQTL coefficients. Having established the predictive ability of peaBrain, we were interested in whether we can use the best-performing peaBrain models to measure the impact of single variants, compared to DeepSEA logFC estimates and MPRA log skew estimates (**Task F**). For all "captured" genes (n=113), we selected all variants identified as significant eQTLs in the GTEx v6p univariate eQTL analysis (n=16,019 variants; **see Online Methods**) and replicated the analysis with the Geuvadis dataset[1] (n=17,279 variants for the EU population and n = 1601 variants for the YRI [Yoruba from Ibadan, Nigeria] population). Unlike GTEx v7, eQTLs from GTEx v6p were derived from genotyping arrays (Illumina OMNI 5M or 2.5M) and thus did not include WGS used to train the peaBrain model. The Geuvdais dataset (both EU and YRI populations) were eQTLs derived from WGS from a set of non-overlapping subjects. For each eQTL (including indels), we created pairs of artificial sequences that only differed at the corresponding snp/indel position and predicted the difference in expression between the alternate and reference alleles from the difference between the two artificial sequences. peaBrain predictions significantly and positively correlated with the univariate eQTL coefficients from the GTEx analysis (Spearman's rho = 0.09; $p = 3.02 \times 10^{-32}$; **Supplementary Figure 2**), from the EU-Geuvadis analysis (rho = 0.10; $p = 9.60 \times 10^{-38}$; **Supplementary Figure 3**), and from the YRI-Geuvadis analysis (rho = 0.18; $p = 8.64 \times 10^{-13}$; **Supplementary Figure 4**). Results were consistent across all datasets when we relaxed our heritability to include genes with GCTA $p < 0.05$: Spearman's rho equal to 0.08 ($p = 3.02 \times 10^{-25}$), 0.07 ($p = 3.65 \times 10^{-25}$), and 0.18 ($p = 1.68 \times 10^{-13}$) for the GTEx, EU-Geuvadis, and YRI-Geuvdais univariately-significant eQTLs, respectively. Alongside this positive correlation, we observed that many variants with large coefficients across all three datasets had small peaBrain predictions (**Supplementary Figures 2-4**). This shrinkage in peaBrain estimates is consistent with the appreciable LD between eQTL-associated variants at any given locus, such that only a subset of significantly-associated variants are actually functional. Whilst univariate linear models are not capable of distinguishing between functional versus "hitchhiker" variants, the joint modelling of the input sequence in Stage 2 peaBrain allows a more direct assessment of the function of each variant. In

comparison, we noted that the MPRA log skew estimates did not correlate with the univariate eQTL coefficients in the GTEx (rho = 0.014; p = 0.60), the EU-Geuvadis (rho = -0.018; p = 0.58), or the YRI-Geuvadis (rho = 0.011; p = 0.82) eQTL analyses. The BiT-STARR-seq log skew estimates also did not correlate with the univariate eQTL coefficients in the GTEx (rho = -0.031; p = 0.50), the EU-Geuvadis (rho = 0.035; p = 0.44), or the YRI-Geuvadis (rho = -0.112; p = 0.32) eQTL analyses. Similarly, for GM12878 (LCL) annotations, we noted that the maximum log fold changes from DeepSEA did not correlate with the magnitude of the eQTL coefficients in the GTEx analysis (rho = -0.001; p = 0.92), the EU-Geuvadis analysis (rho = -0.010; p = 0.18), and the YRI-Geuvadis (rho = 0.016; p = 0.52). The DeepSEA results suggest that simply predicting chromatin effects and TFBS is not sufficient for predicting the transcriptomic consequences of sequence variation for this set of selected genes. This is consistent with experimental evidence suggesting that most genetic variants in DNase footprinted (and other annotated) regulatory regions are silent[2], and the fact that histone modifications (e.g. methylation) alone are frequently insufficient to transcriptionally perturb promoters[3]. However, it is important to note that DeepSEA was trained on the reference genome (i.e. not exposed to genotype variation), while Stage 2 peaBrain was trained to predict differences in expression as a function of differences in sequence between any pair of individuals.

REFERENCES

1 Brown, A. A. *et al.* Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. *Nature Genetics*, doi:10.1038/ng.3979 (2017).

2 Moyerbrailean, G. A. *et al.* Which genetics variants in DNase-Seq footprints are more likely to alter binding? *PLoS genetics* **12**, e1005875 (2016).

3 Ford, E. E. *et al.* Frequent lack of repressive capacity of promoter DNA methylation identified through genome-wide epigenomic manipulation. *bioRxiv*, 170506 (2017).